DOCUMENT RESUME

ED 449 234 TM 032 380

AUTHOR	Ediger, Marlow
TITLE	Assessment and High Stakes Testing.
PUB DATE	2001-00-00
NOTE	9p.
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Educational Assessment; High School Students; High Schools; *High Stakes Tests; *Portfolio Assessment; Portfolios (Background Materials); *Standardized Tests; State Programs; Testing Programs

ABSTRACT

This paper discusses the rationale behind and uses of high stakes testing. Much of the support for high stakes testing comes from a desire to boost student achievement by increasing student and teacher motivation to do well. There are many disadvantages to high stakes testing, including inappropriate use of the test, inadequate test development, and the adverse effects of pressure to perform on students. High stakes tests may not measure what students really need to know or what they are really learning. In many states, schools have not aligned their curricula with the high stakes state-mandated test, and many tests have not been developed to take aspects of the curriculum into account. Students should be evaluated through multiple means, which could include the use of portfolios. Some suggestions are given for portfolio use for learning and assessment. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 449 234

Assessment and High Stakes Testing

Marlow Ediger

TM032380

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Ediger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

ASSESSMENT AND HIGH STAKES TESTING

Much is written about high stakes testing in educational literature. The risks then are great if a student does not pass a state mandated test. A high school senior needs to pass an exit test if he/she is to receive a diploma. Without a high school diploma, the future looks quite bleak for those falling the exit test.

High stakes testing may even include passing a state mandated test to be promoted form one grade level to the next, i.e. grade four to grade five. What is the rationale back of high stakes testing and what are the weaknesses of this plan/policy?

Rationale for High Stakes Testing

Politicians on the state level believe that students need to achieve at a higher rate so that the nation remains competitive among the different countries of the world. The results from the Third International Mathematics and Science Study (TIMMS) has struck fears in the hearts of some and used politically by others to emphasize higher standards and expectations for secondary students to realize. To be sure, there are students who do not apply themselves in course work taken in high school. (See The National Commission on Mathematics and Science Teaching for the 21 st Century, 2000). This was true to some extent when the author attended high school 1942-1946. He still can see in his mind selected secondary students who failed to turn in assigned work on time or not at all. However, one must say also that there were high achievers in all academic disciplines taken. In supervising student teachers for thirty years, the author spoke with many teachers who gave up assigning home work because there was embarrassment on the teacher's part when three out of thirty students in one class turned in their completed home work. As always, there were students, too, who did very well in subject matter areas encountered. It was quite obvious that selected students did very little in class to achieve, but rather attempted to disrupt the class and secure the attention of others to entertain. Three teachers from the same high school in the author's Theory of Administration graduate class openly stated that their school principal was really an excellent person but when sending disciple referrals to him, the misbehavior of the returning pupils was worse than before. These graduate students felt the entire high school was in disarray, even though the same school principal continually invited teachers to send discipline problems to him!

There is little doubt that all individuals could achieve at a higher degree than is presently done. There are also high school students who try very hard and excel in achievement. Others do minimally satisfactory work, and still others could but will not achieve. Mainstreamed students

may/will not be able to attain what is a passing test score on an exit test. This would especially be true of mentally retarded individuals.

A major assumption back of high stakes testing is that teachers are not putting forth the effort to teach well. With pressure put on teachers, they will do better in teaching. Pressures such as school bankruptcy laws, vouchers, merit pay, and charter schools, are there or proposed, as "motivators" for teacher achievement in teaching students. How should teachers be punished for low achieving students living in poverty areas? Students achieve at a much lower rate in poverty areas as compared to suburban region of any sized city (See Ediger, 1996, 34-40).

Disadvantages in High Stakes Testing

Educators tend to be against high stakes testing. There are a plethora of reasons given for this position. First, one test should not determine a students future, such as passing an exit examination to obtain a high school diploma. A question which arises here pertains to the worth of the every day completed school work of students and its value. Should one exit test score be worth more than four years of daily achievement in high school?

Second, are these high stakes tests developed carefully to check validity and reliability in pilot studies? Hastily developed tests may measure little of what it purports to measure. If a test lacks validity, it does not cover what students have had opportunities to learn. Low reliability means the test does not measure consistently, such as alternative forms, split half, an/or test/retest reliability. I f a test does not measure consistently, the results will be flawed. A high standard of error, with statistical data, reveals that there are many weaknesses inherent within that test.

Third, should students have chances to take the test over again if failure was present initially in the first administration of the high stakes test? Certainly, students should be prepared extremely well in order to be successful on the second administration of the exit exam. Otherwise, the dropout rate will be high from the secondary level of schooling. Remedial classes could be conducted in summer sessions, after school hours, or even with special classes during school hours.

Fourth, what is the correlation of the high stakes exit test result with later success at the work place? The test may or may not then measure what is salient for students to learn if the correlation is low. An adequate number of carefully designed studies need to be made in this area. The quality of these studies should be determined by a true panel of experts.

Fifth, should multiple kinds of assessments be used, other than one exit test only, to determine a student's future? Multiple Intelligences

Theory (See Gardner, 1993) has cast doubt on using verbal intelligence only, as is true of reading and responding to multiple choice test items on an exit or any state mandated test. Gardner lists eight intelligences whereby students may reveal what has been learned.

Sixth, what happens to course achievement in those areas not measured by the high stakes test. The slogan, "What gets tested is what gets taught," mesons that non tested academic disciplines may receive short shrift. Many news reports have centered on education whereby teachers drill students so that the latter do well on mandated tests. Is there a way to stress adequately, in teaching and learning situations, untested salient academic disciplines in the school curriculum?

Seventh, how does the worry over passing high stakes tests affect the individual learner? There certainly is much anxiety and tension pertaining to passing a highly important exit test. How many gifted talented students even shudder over failing a high stakes test? Here, the author has to think of a graduate student in his class during the middle 1960s. The high achievement graduate student feared failing the comprehensive, which all needed to pass to receive the Master's Degree. She stated that it was difficult for her to concentrate on her studies due to extreme anxiety in passing the upcoming comprehensive. Later, the same student received her Doctorate in Education (Ed D) and became full professor at a leading university in Illinois. Thus, students may greatly fear in high stakes testing in not receiving a high school diploma. Extreme tension may certainly interfere with the mental status of the individual.

Eighth, do multiple choice test items reveal what a student has learned? These kinds of tests can be machine scored in mass numbers in a matter of minutes. Convenience in scoring is a "virtue." But, does a single test reveal what a student has learned? Responding to multiple choice test items is not like the world of work. Individuals at the work place are not assessed in competency by using a series of multiple choice test items. Rather, they are assessed in terms of how well they do their work, in its totality.

Ninth, can a single numeral state what a student has learned? If a student passes the exit high stakes test, what does this pass/no pass result truly mean in the totality of life itself? What if the student is just barely under the no pass test result, is the standard error of measurement considered in considering the pass/no pass category? The less reliable the test is in measuring consistency of results, the greater the standard error of measurement will be, meaning the test then measures less consistently if the learner took the test over again.

Tenth, how can daily work and achievement of students in class be entered into the pass/no pass philosophy of high stakes testing? In other words, are there other products/processes of students which should provide some kind of average of achievement, other than the test

score itself? (See Popham (1999).

Test writers and policy makers need to take into consideration the above ten named concerns and attempt to remedy the weaknesses. Tests need to be of much higher quality than what is now available. Test results may then become more important to use for diagnostic and remediation purposes. Problems of test results from high stakes testing are numerous. The following are quite apparent from Education Week (January 24, 2001):

1. North Carolina delayed by one school year implementation of their high stakes testing plan. Alaska delays implementation by four years from 2002 to 2006; whereas Arizona delayed their implementation until 2004, due to initial high failure rates. Wyoming will us a body evidence to notice if students have met the highest of three levels for high school graduation, the proficiency level

2. weaknesses in tests being too difficult for students to achieve well in has prompted California to take the difficult items out of mathematics tests and shorten the high school exit exam by one hour. Ohio has done away with a required test for fourth graders to pass in order to be promoted to fifth grade. In Louisiana, an 8th grade student may take the state test three times in order to succeed. If failure to pass the state mandated test is still there, the student may receive an alternative program leading to a GED.

3. in Massachusetts, 67% passed the English section of high stakes test while 44% passed the mathematics test in 1999. By 2003, students will need to pass the English and mathematics section of the state mandated test to receive a high school diploma. The Board of Education wondered if the the questions were too difficult.

According to the above, as well as other content, in the above named article, it appears that

1. many states set standards too high for students to graduate from high school, resulting in a high rate of failure.

2. many schools did not have the opportunity to align their curriculum with the state mandate standards of instruction. Validity in testing is then lacking.

3. many states did not try their tests out in pilot studies to take out the kinks and weaknesses inherent. A large standard error of measurement will then be in evidence.

4. some states knew their tests were too difficult; they took out the excessively difficult test items prior to administering the test.

5. high stakes testing within grade levels, such as from grade four to grade five, are not perceived as having any merit.

Role of Students in High Stakes Testing

Students may worry much about passing a high stakes test. This may be true of all ability levels of students. The toll here may also be high in terms of students being distracted from the every day learning activities and experiences provided in the classroom. The student may come to feel that it does not matter much what was done in the classroom; it is the single exit test score which counts.

Multiple choice test items tend to be isolated from each other. Students should perceive the relationship of ideas rather then isolated bits of subject matter. Testing might set up unrealistic situations to ascertain what students have learned. Learning should be in context.No other group takes tests to ascertain what is being done well, except teachers with competency tests and students with high stakes and other tests to measure academic achievement. Workers at the work place are assessed in terms of life like competencies possessed on the job or profession, not through the taking of tests.

The National Council Teachers of English (NCTE) passed a Resolution on Developing a Test Takers Bill of Rights including the following:

1. the right that standardized tests be adopted through an open, public process that considers the design and appropriateness of the test;
2. the right to know before the test date the form of any given test;
3. the right to know how the results of the test will be used;
4. the right to a challenging curriculum that is not constrained by any given test;
5. the right to arrange accommodations for documenting learning differences and/or unforeseeable circumstances;
6. the right to display competencies through various means.;
7. the right to an open process of review of test items and results;
8. the right to challenge test scores and have them changed if they are incorrect;
9. the right to a process that corrects tests and /or individual items found to be invalid or unreliable (See Iowa Language News, 2000).

The author would include the right of parents/students to challenge machine scored results if computer glitches were expected in any student's test results. Students and parents do have rights in a hierarchical means of ascertaining the former's progress.

The Portfolio, an Alternative Means of Assessment

Multiple means of evaluating student achievement could incorporate the portfolio philosophy of assessment. Portfolios place

major emphasis upon what the student achieves on a daily basis. What is achieved in the regular classroom in terms of activities and experiences then become salient, not a single test score. The student and the teacher cooperatively may develop a portfolio for the former to reveal achievement and progress. Contents of the portfolio need to be selected carefully. there are many products which may go into a portfolio from the daily work of the learner. A portfolio then might become too voluminous for assessors to appraise. A balance needs to be in the offing whereby a portfolio is not too voluminous as compared to being too thin in entries. With the latter, an assessor might not be able to get a total picture of the learner's achievement.

What may be chosen in terms of daily student work to go into a portfolio?

1. written work including poems, stories, summaries, outlines, and conclusions relating directly to an ongoing lesson/unit of study.
2. art products including murals, pencil sketching, water color paintings, friezes, bulletin board displays, and collages which relate to objectives of instruction.
3. snapshots of construction items, too large to include in a portfolio, such as of a model, an object, a display, a diorama, background scenery for a play, as well as a set of puppets and marionettes.
4. oral communication, individual or within a committee, including book reports, extemporaneous speeches, dramatic experiences, discussions, collaborative endeavors, oral reading from the basal textbook, making introductions, planning a classroom exhibit, a readers' theater presentation, story telling, singing, reporting, choric reading, explanations, telling jokes, oral communication club activities, rope jumping chants, tongue twisters, finger plays, flannel board talks, and peer oral experiences involving achievement of vital objectives of instruction.
5. rubric results on listening activities including self assessment using appropriate criteria, teacher evaluation in terms of objectives, peer evaluation of being a good listener, student results from a standardized listening test, and committee members assessment of listening progress (an open ended approach) within an ongoing learning activity (Ediger, 2000, 503-505).

A Table of Contents should be developed by the involved student for the portfolio. The purpose of the portfolio should be carefully written. It is best if professionally trained individuals assess portfolio contents. If teachers are to do the assessing, adequate compensation should be in the offing. Raters of the portfolio need to study the criteria in the rubric for use in the assessment process. Interscorer reliability is important. Thus for any one portfolio, raters need to agree upon the rating given on

a five point Likert scale. If there is a lack of agreement, a negotiated rating may then be given. There will be more subjectivity involved in rating portfolios as compared to test results. It is impossible to give a single numeral, as is the case of test results, when indicating how well a student did on portfolio contents. However, the contents in a portfolio pertain to every day work of students, not on isolated test items in a multiple choice test written by those far removed from the local classroom. Multiple choice test items are taken out of context whereas portfolio contents pertain to authentic work of the student. An observant teacher may then notice sequential progress of the learner in one area, for example, such as written work. Earlier written work may then be compared with later products of writing of the individual student.

References

Ediger, Marlow (1996), "Evaluation of Pupil Achievement," Education Magazine, nr. 118, pp 34-40. Published by the Qatar National Commission for Education -- the Middle East.

Ediger, Marlow (2000), "Choosing Evaluation Procedures," Education, 120 (3), 503-505.

Education Week (January 24, 2001), "States Adjust High Stakes Testing Plans," pp.1, 18.

Gardner, Howard (1993), Multiple Intelligences: Theory Into Practice. New York: Basic Books.

Iowa Language News (Winter, 2000), Scott Romine, Editor, ""English Teachers Pass Resolution on High Stakes Testing and the Rights of Test Takers." Published by the Iowa Council Teachers of English.

Popham, W. James (1999), "Why Standardized Tests Do Not Measure Educaitonal Quality," Educational Leadership 56 (6), 8-15.

The Nationl Commisssion on Mathematics and Science TEaching for the 21st Century (2000), Before It's Too Late. Jessup, Maryland: Eduation Publications Center, U. S. Departmetn of Educaiton.